UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/053,884	01/22/2002	Eetu Ojanen	540-015.2	5315

			EXAMINER
4955 7590 09/28/2004			PANNALA, SATHYANARAYA R

WARE FRESSOLA VAN DER SLUYS &
ADOLPHSON, LLP
BRADFORD GREEN BUILDING 5
755 MAIN STREET, P O BOX 224
MONROE, CT 06468

ART UNIT	PAPER NUMBER
2177	11

DATE MAILED: 09/28/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/053,884	OJANEN, EETU	
	Examiner	Art Unit	
	Sathyanarayan Pannala	2177	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) FROM
THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed
  after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any
  earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *22 January 2002*.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is
    closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-24* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-24* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☒ All   b)☐ Some * c)☐ None of:

    1.☒ Certified copies of the priority documents have been received.

    2.☐ Certified copies of the priority documents have been received in Application No. _____.

    3.☐ Copies of the certified copies of the priority documents have been received in this National Stage
        application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
    Paper No(s)/Mail Date *1 & 8*.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: _____.

DETAILED ACTION

1.   The Application 10/051277 filed on 1/22/2002 has been examined. Claims 1-24 are pending in this Office Action.

Priority

2.   Acknowledgment is made of applicant's claim for foreign priority under 35 U.S.C. 119(a)-(d). The certified copies of priority filed with the Application No. 10/053884 on 1/22/2002 have been received.

Claim Rejections - 35 USC § 101

3.   35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of

matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the

conditions and requirements of this title.

4.   Claims 1-22 are rejected under 35 U.S.C. 101, because independent claim 1 is directed to a method for processing information and claim 15 is directed to an arrangement for processing information, whereas claim 22 is directed to a computer program code, which are all non-statutory subject matter.

As per independent claims 8 and 15 the preamble recites "A method and an arrangement" as drafted said claim is not technologically embodied to a computer, whereas the independent claim 22, the preamble recites a computer program code" as drafted said claim is not an utility embodied to computer(See *In re Waldbaum*, 173 USPQ 430 (CCPA 1972); *In re Musgrave*, 167 USPQ 280 (CCPA 1970) and *In re Johnston*, 183 USPQ 172 (CCPA 1974) also see MPEP 2106 IV 2(b), even though said claim is limited to a useful, concrete and tangible application (See *State Street v. Signature financial Group*, 149 F.3d at 1374-75, 47 USPQ 2nd at 1602 (Fed Cir. 1998); AT&T Corp. V. Excel, 50 USPQ 2nd 1447, 1452 (Fed. Cir. 1999).

Claim Rejections - 35 USC § 102

5.   The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

6.   Claims 1-24 are rejected under 35 U.S.C. 102(b) as being anticipated by Ribeiro-Neto et al, (ACM Publisher, "Extracting Semi-Structured Data Through Examples", 1999) hereinafter Neto.

7.   As per independent claim 1, Neto anticipated by teaching an approach to

extracting semi-structured data from Web sources by collecting a couple of example

objects from the user and this information is used to extract new objects from new

pages or texts (page 94, paragraph Abstract). Neto teaches the claimed step of

"pointing out at least two exemplary cases" as a couple of examples are sufficient for

extracting hundreds of objects from new web pages (page 94, col. Right, paragraph

one). Further, Neto teaches the claimed step of "comparing the at least two exemplary

cases to each other for finding congruent parts from them" as we investigate (compare)

how to extract objects and their attributes to insert into nested tables for later querying.

Using the user examples a strategy is devised to extract data form similar structure

(page 95, col. Left, paragraph last and col. Right, paragraph 2). Further, Neto teaches

the claimed step of "as a result of the comparing, generating a regular expression,

which describes the appearance of congruent parts in the at least two exemplary cases"

as by properly extracting objects and their attributes and inserted into tables for later

querying (page 95, col. Left, paragraph last). Further, Neto teaches the claimed step of

"on the basis of the generated regular expression, generating a set of rules for

extracting data of a desired kind" as once an object is properly structured it can be

directed inserted into a nested tables for later querying (Fig. 3, page 95, col. Right,

paragraph last). Finally, Neto teaches the claimed step of "extracting data areas from

the original data according to the generated set of rules" as for each piece of data in the

example object in the figure we assume that we know the position in the original page

where it came from (Fig. 2, page 95, col. Left, paragraph last).

8.   As per dependent claim 2, Neto teaches the claimed step of "comprising the step of modifying the extracted data areas to be uniform in format" as the nested table can be flattened for querying as a standard relational table (Fig. 3, page 95, col. Left, paragraph last).

9.   As per dependent claim 3, Neto teaches the claimed step of "the at least two exemplary cases that are pointed out each have a structure and a content, and the structure of the exemplary cases is identical, but the content is different" as given a very small set of example objects, there are several possible strategies for extracting data from new pages with similar structure. The given algorithm works by assembling a context for an object (page 96, col. Left, paragraph last).

10.   As per dependent claim 4, Further, Neto teaches the claimed step of " the exemplary cases are pointed out from the original data" as for each piece of data in the example object, the assumption is that the position of is known in the original page where it came from (Fig. 2, page 95, col. Right, paragraph last).

11.   As per dependent claim 5, Neto teaches the claimed step of "the regular expression comprises the congruent parts and wildcard expressions, which correspond to matter to be extracted" as attribute value pair is shown and the symbol "*" (wildcard) matches a sequence of characters of any length (page 95, col. Left, paragraph last).

12.   As per dependent claim 6, Neto teaches the claimed step of "the set of rules generated on the basis of the regular expression is stored for further usage" the extracted objects (rules) are stored in a regular text files using a XML based format that allows for easy conversion to other formats or for insertion in nested tables for later querying (page 97, col. Left, paragraph first).

13.   As per dependent claim 7, Neto teaches the claimed step of "the step of tokenizing the chosen exemplary cases prior to the processing proper thereof, by replacing certain elements of the exemplary cases by corresponding data structures, which contain an identifier, such as a type characteristic, or a name, as well as a data content of said element" as given an AVP selected by the user, then determine a passage surrounding the AVP value in the text (page 97, col. Right, paragraph 2-3).

14.   As per dependent claim 8, Neto teaches the claimed step of "between the at least two exemplary cases pointed out, there is at least one identical element, the counterpart whereof in the treatment of the exemplary cases is a given token" as to extract information form a set of data rich pages the assumption of the existence of a grammar detailing how to parse and recognize tokens for insertion on a table (page 95, col. Right, paragraph first).

15.   As per dependent claim 9, Neto teaches the claimed step of "in order to generate

a set of rules, the method comprises the steps of: marking the longest of the selected,

tokenized examples as a regular expression, marking the next longest of the selected,

tokenized examples as an exemplary expression and comparing the regular expression

with the exemplary expression of the moment in question" as object extraction patterns

are used by the extractor module to find and extract new objects form web pages

(Fig. 4-5, 7, page 97, col. Right, paragraphs in section 4.2).

16.   As per dependent claim 10, Neto teaches the claimed step of "the regular

expression and the exemplary expression of the moment in question are compared by

means of a given reference algorithm that returns an edit script" as the bottom-up

extraction is that it recognizes and extracts atomic object components prior to

recognition of the object itself and the component objects are used to assemble the

object through a bottom-up composition operation (Fig. 8, page 98, col. Left, paragraph

first in section 5.2).

17.   As per dependent claim 11, Neto teaches the claimed step of "the regular

expression and the exemplary expression of the moment are compared by means of a

reference algorithm that returns the shortest possible edit script" as for each AVP

pattern, get all strings and store them in AVP_BAG along with positional information

(Fig. 8, page 98, col. Left, paragraph first in section 5.2).

18.   As per dependent claim 12, Neto teaches the claimed step of "in order to

generate a set of rules, the regular expression is modified according to the edit

information contained in the edit script" as the objects being composed might result for

not including all components of OE pattern (Fig. 8, page 98, col. Left, paragraph last to

col. Right, paragraph first).

19.   As per dependent claim 13, Neto teaches the claimed step of "the created

regular expression constitutes a set of rules by itself" as the list is formed by several

identical complex objects and each of these objects is composed of a list and two atoms

and the author list is itself formed by atomic objects (page 98, col. Left, paragraph

second).

20.   As per dependent claim 14, Neto teaches the claimed step of "by means of the

generated set of rules, from the original data there are extracted elements according to

the exemplary cases" as the GUI provides the user with a java interface to assemble a

couple of examples and assembled objects are then used to generate patterns for

extracting new objects (Fig. 5, page 96, col. Right, paragraph third).

21.   As per independent claim 15, Neto anticipated by teaching an approach to

extracting semi-structured data from Web sources by collecting a couple of example

objects from the user and this information is used to extract new objects from new

pages or texts (page 94, paragraph Abstract). Neto teaches the claimed step of

"pointing out at least two exemplary cases, means for comparing the at least two

exemplary cases to each other for finding congruent parts from them" as a couple of

examples are sufficient for extracting hundreds of objects from new web pages (page

94, col. Right, paragraph one). Further, Neto teaches the claimed step of "generating,

as a result of the comparing, a regular expression, which describes the appearance of

congruent parts in the at least two exemplary cases" as by properly extracting objects

and their attributes to insert into nested tables for later querying. Using the user

examples a strategy is devised to extract data form similar structure (page 95, col. Left,

paragraph last and col. Right, paragraph 2). Further, Neto teaches the claimed step of

"generating a set of rules on the basis of the generated regular expression, in order to

extract desired information" as once an object is properly structured it can be directed

inserted into a nested tables for later querying (Fig. 3, page 95, col. Right, paragraph

last). Finally, Neto teaches the claimed step of "extracting data areas from the original

data according to the generated rules" as for each piece of data in the example object in

the figure we assume that we know the position in the original page where it came from

(Fig. 2, page 95, col. Left,

paragraph last).

22.   As per dependent claim 16, Neto teaches the claimed step of "modifying the

extracted elements to be uniform in format" as when used the local context information

is very specific and therefore it would not retrieve any author name other than Eric

Simon. Thus to be able to generate AVP pattern for extracting other author names a

more flexible pattern generation strategy is used (Fig. 6, page 97, col. Right,

paragraph one).

23.   As per dependent claim 17, Neto teaches the claimed step of "in order to point

out exemplary cases, the arrangement is provided with pointers to character strings" as

the example object with a hierarchical structure the web page (Fig. 1-3, page 95, col.

Right, paragraph last).

24.   As per dependent claim 18, Neto teaches the claimed step of "tokenizing the

examples pointed out by replacing given elements of the exemplary cases by

corresponding data structures that contain a type characteristic or a name as well as a

data content of said element" as given an AVP selected by the user, then determine a

passage surrounding the AVP value in the text (page 97, col. Right, paragraph 2-3).

25.   As per dependent claim 19, Neto teaches the claimed step of "processing

tokenized data" as to extract information form a set of data rich pages the assumption of

the existence of a grammar detailing how to parse and recognize tokens for insertion on

a table (page 95, col. Right, paragraph first).

26.   As per dependent claim 20, Neto teaches the claimed step of "generating a set of

rules according to a created regular expression" as an AVP selected by the user then

determine a passage surrounding this AV value in the text and the symmetric passage

composed of width (W) text tokens to the right and W text tokens to the left of the AVP

value (Fig. 6, page 97, col. Right, paragraphs 2-3).

27.   As per dependent claim 21, Neto teaches the claimed step of "generating the set

of rules including a program component created especially for this purpose, which

program component is different from the program component that is meant for

extracting data areas by using the generated set of rules" as the simple algorithm works

by assembling a context for an object and using this context description to identify new

objects in new pages, example-based approach requires an environment which allows

the specification of examples and the extraction of the semi-structured data and it is

called as data extraction by example (DEByE). The extractor mode takes the generated

patterns and applies them to new pages form the target web pages (Fig. 4, 7-8,

page 96, col. Left, paragraph last and col. Right, paragraphs first and last).

28.   As per independent claim 22, Neto anticipated by teaching an approach to

extracting semi-structured data from Web sources by collecting a couple of example

objects from the user and this information is used to extract new objects from new

pages or texts (page 94, paragraph Abstract). Neto teaches the claimed step of

"pointing out at least two exemplary cases" as a couple of examples are sufficient for

extracting hundreds of objects from new web pages (page 94, col. Right, paragraph

one). Further, Neto teaches the claimed step of "comparing the at least two exemplary

cases to each other for finding congruent parts from them" as we investigate (compare)

how to extract objects and their attributes to insert into nested tables for later querying.

Using the user examples a strategy is devised to extract data form similar structure

(page 95, col. Left, paragraph last and col. Right, paragraph 2). Further, Neto teaches

the claimed step of "as a result of the comparing, generating a regular expression,

which describes the appearance of congruent parts in the at least two exemplary cases"

as by properly extracting objects and their attributes and inserted into tables for later

querying (page 95, col. Left, paragraph last). Further, Neto teaches the claimed step of

"on the basis of the generated regular expression, generating a set of rules for

extracting data of a desired kind" as once an object is properly structured it can be

directed inserted into a nested tables for later querying (Fig. 3, page 95, col. Right,

paragraph last). Finally, Neto teaches the claimed step of "extracting data areas from

the original data according to the generated set of rules" as for each piece of data in the

example object in the figure we assume that we know the position in the original page

where it came from (Fig. 2, page 95, col. Left, paragraph last).

29.   As per independent claim 23, Neto anticipated by teaching an approach to

extracting semi-structured data from Web sources by collecting a couple of example

objects from the user and this information is used to extract new objects from new

pages or texts (page 94, paragraph Abstract). Neto teaches the claimed "pointing out at

least two exemplary cases" as a couple of examples are sufficient for extracting

hundreds of objects from new web pages (page 94, col. Right, paragraph one). Further,

Neto teaches the claimed "comparing the at least two exemplary cases to each other for

finding congruent parts from them" as we investigate (compare) how to extract objects

and their attributes to insert into nested tables for later querying. Using the user

examples a strategy is devised to extract data form similar structure (page 95, col. Left,

paragraph last and col. Right, paragraph 2). Further, Neto teaches the claimed step of

"as a result of the comparing, generating a regular expression, which describes the

appearance of congruent parts in the at least two exemplary cases" as by properly

extracting objects and their attributes and inserted into tables for later querying (page

95, col. Left, paragraph last). Further, Neto teaches the claimed step of "on the basis of

the generated regular expression, generating a set of rules for extracting data of a

desired kind" as once an object is properly structured it can be directed inserted into a

nested tables for later querying (Fig. 3, page 95, col. Right, paragraph last). Finally,

Neto teaches the claimed "extracting data areas from the original data according to the

generated set of rules" as for each piece of data in the example object in the figure we

assume that we know the position in the original page where it came from (Fig. 2, page

95, col. Left, paragraph last).

30.   As per independent claim 24, Neto anticipated by teaching an approach to

extracting semi-structured data from Web sources by collecting a couple of example

objects from the user and this information is used to extract new objects from new

pages or texts (page 94, paragraph Abstract). Neto teaches the claimed step of

"causing a computer to execute a procedure that comprises the steps of: pointing out at

least two exemplary cases" as a couple of examples are sufficient for extracting

hundreds of objects from new web pages (page 94, col. Right, paragraph one). Further,

Neto teaches the claimed step of "comparing the at least two exemplary cases to each

other for finding congruent parts from them" as we investigate (compare) how to extract

objects and their attributes to insert into nested tables for later querying. Using the user

examples a strategy is devised to extract data form similar structure (page 95, col. Left,

paragraph last and col. Right, paragraph 2). Further, Neto teaches the claimed step of

"as a result of the comparing, generating a regular expression, which describes the

appearance of congruent parts in the at least two exemplary cases" as by properly

extracting objects and their attributes and inserted into tables for later querying (page

95, col. Left, paragraph last). Further, Neto teaches the claimed step of "on the basis of

the generated regular expression, generating a set of rules for extracting data of a

desired kind" as once an object is properly structured it can be directed inserted into a

nested tables for later querying (Fig. 3, page 95, col. Right, paragraph last). Finally,

Neto teaches the claimed step of "extracting data areas from the original data according

to the generated set of rules" as for each piece of data in the example object in the

figure we assume that we know the position in the original page where it came from

(Fig. 2, page 95, col. Left, paragraph last).

Conclusion

Any inquiry concerning this communication or earlier communications from the
examiner should be directed to Sathyanarayan Pannala whose telephone number is
(703) 305-3390. The examiner can normally be reached on 8:00 am - 5:00 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's
supervisor, John Breene can be reached on (703) 305-9790. The fax phone number for
the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the
Patent Application Information Retrieval (PAIR) system. Status information for
published applications may be obtained from either Private PAIR or Public PAIR.
Status information for unpublished applications is available through Private PAIR only.
For more information about the PAIR system, see http://pair-direct.uspto.gov. Should
you have questions on access to the Private PAIR system, contact the Electronic
Business Center (EBC) at 866-217-9197 (toll-free).

Sathyanarayan Pannala
Examiner
Art Unit 2177

srp
September 19, 2004